DOCUMENT RESUME

ED 437 398 TM 030 550

AUTHOR Farrell, Donna M.; Hanson, Ralph A.

TITLE The Golden Rules for Assuring Gains in Student Learning from School Programs.

PUB DATE 1999-11-00

NOTE 14p.; Paper presented at the Annual Meeting of the California Educational Research Association (Monterey, CA, November 17-19, 1999).

PUB TYPE Reports - Descriptive (141) -- Speeches/Meeting Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.

DESCRIPTORS *Academic Achievement; *Achievement Gains; *Educational Improvement; Elementary Secondary Education; *Evaluation Methods; Formative Evaluation; Learning; *Program Evaluation; Research Utilization; Summative Evaluation

ABSTRACT

This paper uses research findings on program evaluations over the last 20 years to derive 10 "Golden Rules" that will help school districts increase student learning. Each of the rules is described and illustrated within the context of a five-part quality assurance and evaluation system of program screening, field-testing, adoption and implementation, summative evaluation, and confirmative evaluation. The rules are: (1) identify programs that are closely aligned with state and district standards; (2) identify programs that have some evidence of established validity; (3) field test all programs considered for district adoption at least 2 years before their full adoption; (4) make sure there is adequate teacher training and staff development for implementing all program components; (5) make program adoption decisions based primarily on field study results; (6) tailor the adopted program to fully comply with district requirements during the first year of implementation and purchase any needed supplemental instructional materials; (7) design an instructional verification system to accompany implementation and record the amount of instruction provided to each class, as well as each class' level of proficiency on the program outcome ; (8) review and analyze program cost, implementation, and effects data at least three times a year and develop intervention strategies to assist in problem areas; (9) produce a summative evaluation report each year; and (10) produce a confirmative evaluation report after 3 years of program implementation. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

SCOPE OF INTEREST NOTICE

The ERIC Facility has assigned this document for processing to: TM

In our judgment, this document is also of interest to the Clearinghouses noted to the right. Indexing should reflect their special points of view. EA

ED 437 398

[1]The Golden Rules for Assuring Gains in Student Learning From School Programs

Donna M. Farrell, Coordinator of Research and Evaluation
Moreno Valley Unified School District
and
Ralph A. Hanson, Educational Measurement Specialist
Hanson Research Systems

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

D. Farrell

1

[1]Paper presented at the annual conference of the California Educational Research Association, November 17-19, 1999, Monterey, CA. For further information or requests for paper, contact Donna Farrell, (909) 485-5600 Ext. 2235 or e-mail: dfarrell@mvusd.k12.ca.us.

TM030550

BEST COPY AVAILABLE

The Golden Rules for Assuring Gains in Student Learning from School Programs

Donna M. Farrell, Moreno Valley Unified School District
Ralph A. Hanson, Hanson Research Systems

George Bernard Shaw once said, "The golden rule is, there is no golden rule." However, in the day-to-day world of school program evaluation, a temporary dispensation from the ultimate truth of this axiom is required. This paper uses the research findings on program evaluations during the past 20 years to derive ten "Golden Rules" that will help school districts to increase their students' learning.

Each of the ten golden rules are described and illustrated in detail within the context of a five-part quality assurance and evaluation system: program screening, field-testing, adoption/Implementation, summative evaluation, and confirmative evaluation (See Figure 1, page 11). The activities carried out within each part of this research and development (R & D) model of program evaluation are designed to provide answers to a specific research question and produce a useful report on program costs and benefits. Adhering to these ten golden rules as closely as possible provides a template for school districts to select, Implement, and evaluate valid instructional programs and markedly improve their students' learning.

PART I: SELECTION SCREENS

The rationale for golden rules one and two is to answer the research question: Does the program fit district needs and requirements? The product of this research is a report describing the 'fit' of candidate programs with district requirements and recommendations on which programs to carry out field studies.

Golden Rule Number One

Identify programs that are closely aligned with state and district standards.

Although this first rule appears to be stating the obvious, finding such programs is not an easy task. Most educational programs do not provide **clear and obvious** instruction in the required state and district standards. Further, the relationship of the lesson content, practice activities, and assessments to the identified standard(s) is not always readily apparent. Publishers want to sell their programs to the nation, as well as to individual states; therefore, their programs are designed to develop specific skills, produce particular learner outcomes and/or meet national standards in the content area. This program content is then "aligned" with individual state and district standards.

As a result, there is not always an explicit match between program lessons and state/district standards. In many cases, the instructional materials are neither in

a logical format for day-to-day classroom instruction, nor adequately developed to allow students to meet the state standards. For example, most program materials typically include a matrix of the state/district standards matched to the program lessons. However, a more user-friendly format for teachers is to have a matrix of each program lesson matched to the standard that it teaches. The "inverted" matrix typically provided to teachers requires them to spend unnecessary preparation time looking-up the standards that go along with the lessons they are planning to teach. Then, to add insult to injury, they generally have to seek out or develop additional practice activities and/or assessments to provide them with sufficient data on their students' progress towards meeting the standards.

Golden Rule Number Two

Identify programs that have some evidence of established validity.

Evidence of established validity means that the program has research to show that it has been through a formative evaluation or has been field tested in actual school settings. Ideally, all program developers and publishers should conduct formative evaluations and field studies well in advance of wholesale marketing and use. However, this is rarely the case. The more usual situation is that districts must conduct what is called an evalubility analysis on programs already published (Wholey, 1979).

This requires a careful analysis of all instructional materials and components, including the determination of the time, personnel, and other costs involved. The purpose of this analysis is to determine if the program meets the technical criteria to be considered a valid, researched-based program. To be selected for a field study a program should meet the following criteria:

- Have an architecture that enables teachers and administrators to monitor and evaluate student progress in the program, as well as to verify the amount of program implementation within each classroom (i.e., pre-, post, and benchmark tests carefully aligned to program lessons and practice activities).
- Identify the target population for which it is intended.
- Have a stated purpose, as well as instructional goals, learner outcomes, and an identified duration time in which to achieve the goals and produce the outcomes.
- Include all of the educational materials, activities and experiences to accomplish its goals and produce the learner outcomes with the target population(s) (i.e., teachers are not required to purchase, create, or locate the materials necessary to implement lesson activities).
- Provide practice activities and assessment materials that are carefully aligned with the lesson content and that focus on the program's goals and learner outcomes.

- Provide a training system to train teachers and other users to properly implement and use the program.

Finally, a valid, research-based program should have field studies to verify student learning in the general population, as well as in various sub-populations of interest (e.g., different grade-levels and ethnic groups; basic skills and gifted students; males and females). The field studies should validate the program's stated goals and learner outcomes and provide perspective users with some degree of assurance that, when fully implemented with similar populations, the program will "work."

PART II: DISTRICT FIELD STUDY

The rationale for golden rules three through five is to answer the research question: Will the program work? The product of this research is a report on the field study results with recommendations for which program to adopt.

Golden Rule Number Three

Field-test all programs considered for district adoption AT LEAST 2 years prior to their full adoption.

The vast majority of programs have very limited field study results, if they have any at all. However, even if a program has excellent field study results, it should still should be field-tested within the school district. Based on the evaluablity criteria listed in rule number one, districts should select two or three programs to field-test at least two years prior to the mandatory adoption date.

A field study gives the district an opportunity to identify the specific problems associated with the implementation of the program BEFORE it is adopted. For example, a field study gives the district time to identify the program's areas of weaknesses. Information can also be obtained from teachers on the program's ease of use, and teacher-training requirements can be identified. Finally, data can be collected and analyzed to determine the program's effects on student learning.

The decision as to which program to adopt can then be made on the bases of the program's cost verses its benefits. For example, if one or more of the following typical implementation problems emerge during the field study, the district may choose not to adopt the program:

- There was an insufficient amount of instruction and practice activities on a number of standards,
- Assessment components were poorly aligned with the instructional materials,

- Most teachers expressed dissatisfaction with the amount of time it took to learn and implement the program components,
- The program required an extensive amount of teacher training,
- The program produced only small gains in student achievement.

The cost of correcting any one of theses major design flaws simply may not be worth the benefits the program provides.

Golden Rule Number Four

Make sure there is adequate teacher training and staff development for implementing all program components.

Most teachers are quite satisfied with their newly adopted programs until they are held accountable for implementing them. That is yet another reason field-testing is so important. Teachers need to know how easy it is for them to learn and implement the program and how much training and assistance they will need. Some programs are very straightforward, have few components, and require little teacher training. Others are more comprehensive, have many different components, and require several days or weeks of staff-development.

In the case of comprehensive programs, teachers need to know which components they are expected to implement and how to use them to ensure student learning. For example, most middle and high school literature programs have so many lessons and activities, that it is impossible for teachers to cover all of the material during a typical school year. Based on the analysis of the program during the evaluability analysis, the required district standards, and teachers' 'wisdom of practice', decisions must be made regarding which components will be used, what lessons must be taught and what skills and concepts must be assessed. Teacher training and staff development can then be provided for the identified components.

Golden Rule Number Five

Make the program adoption decisions based primarily on the results of your district's field study.

While it is important to have teacher input and 'buy-in' on program adoption decisions, their contribution to the selection process should come within the context and structure of the evalubility analysis and field-study activities. Teachers can help to analyze program components to determine if they are aligned with district standards and meet the technical criteria listed in rule number one. If their analysis determines that most of the criteria are met, then the program is eligible to be field-tested. During the year of field-testing,

teachers should also provide the district with information documenting the program's ease of use, curriculum alignment, support and training, and overall completeness and effectiveness.

This implementation information, along with student achievement and program cost data, is used by the district to decide which program it should adopt. Bottom line, the decision must be based on which program provides the greatest gains in student learning for the least amount of effort and cost.

PART III: PROGRAM ADOPTION AND IMPLEMENTATION

The rationale for golden rules six through eight is to answer the research question: What will it take to make the program work? The product of this research is a summative evaluation report on the first-year implementation

Golden Rule Number Six

Tailor the adopted program components to fully comply with district requirements during the first year of implementation and purchase/develop any needed supplemental instructional materials.

No matter which program is ultimately selected, some of the components will either be missing altogether or have to be adapted in some manner to comply with the needs of the district. Thus, an important part of the first-year implementation process is: (1) Identifying the supplemental materials that are needed and purchasing them; and (2) Identifying the specific tailoring needs and adapting the components.

One example of program tailoring is adapting the assessment component of a program. Typically the benchmark assessments (e.g., end of chapter/unit tests) that accompany programs are not well aligned with lesson objectives, content, and practice activities. Also, they may not provide sample responses to open-ended items. In addition, many programs do not include pre- and posttests designed to assess students' learning of the program's full instructional content. Finally, rarely are any field studies carried out on the reliability and validity of any of these assessments. As a result, pre- post and benchmark tests must be either developed or adapted to provide teachers and administrators with valid measures of the level of student attainment of lesson objectives and to verify the amount program implementation within classrooms.

Another common example of tailoring is developing additional practice activities and/or purchasing supplemental materials for students to learn the skills and concepts necessary to meet state/district standards. This is usually done because:

- The program does not provide a sufficient amount of practice for students to learn a skill or meet a standard;
- The practice activities are not in the same format or do not use the same vocabulary as the assessment activities; and/or
- The practice activities are not at the appropriate skill level for students.

Golden Rule Number Seven

Design an instructional verification system to accompany implementation and record the amount of instruction provided to EACH class, as well as each class' level of proficiency on the program outcome.

Although implementation factors have long been known to impact student achievement, for the most part, this critical aspect of program evaluation has been ignored. Program evaluation studies rarely obtain any measures of the amount of implementation. The assumption is that all teachers fully implement the district-adopted programs placed in their classrooms exactly as prescribed. However, as most administrators know all too well, this is an extremely erroneous assumption. Teachers implement programs in a variety of ways.

What school administrators need to know is, "Do the differences in teacher implementation practices impact their students' learning?" In other words, do those teachers who implement 'more' of the program have students with higher levels of achievement than those who implement 'less'? And if so, which lessons or program components have the greatest impact?

To gather such data, districts must develop a program implementation verification system to collect data on the amount program implementation in each class at each school. This can be done in several ways. One way is simply to have teachers fill out a short questionnaire, on a regular basis, indicating the scope and sequence of the program content that they covered in their classes.

Another way is to collect the students' scores on the assessments that accompany the program (e.g., end of unit/chapter tests). Although this requires developing a more comprehensive data management system, it can serve a number of important purposes. First and foremost, It helps to ensure that the program will be fully implemented in the classroom (i.e., what gets tested, gets taught). Second, it validates the alignment of the program tests with program lessons and practice activities. Third, it validates student learning; and fourth, It is a valuable tool for teachers in developing lesson plans and providing individualized instruction. The district can then use these data to conduct summative and confirmative (i.e., follow-up) evaluations on the program to validate learner outcomes with district students and its long-term effects.

Golden Rule Number Eight

Review and analyze the program cost, implementation, and effects data at least three times a year and develop intervention strategies to assist in problem areas.

As the program implementation data are collected each quarter or trimester, it should be analyzed to verify that:

- the program is being fully implemented in the appropriate classrooms across the district;
- the anticipated effects are occurring (i.e., student learning) in a variety of contexts; and
- the costs remain as anticipated (e.g., no unexpected costs emerge, such as excessive duplicating expenses and/or having to purchase additional supplemental materials).

Collectively, these data should provide evidence that: 1) the program is valid, reliable and replicable, and 2) any further evaluations carried out on it (i.e., summative and confirmative) will be able to show the logical relationships that exist between program implementation and learner outcomes, as well as between instruction and learning.

PART IV: SUMMATIVE EVALUATION

The rationale for golden rule nine is to answer the research question: How well does the program work? The product of this research is an annual summative evaluation report with program costs and benefit information.

Golden Rule Number Nine

Produce a summative evaluation report each year.

The evalubility analysis, field study research and development efforts, and the Instructional Verification System allow the summative evaluation process to emerge. The focus of a summative evaluation is on measuring program effects. That is, measuring the direct, program-specific effects that are attained by students and identifying the conditions under which they attain them.

The summative evaluation is a dynamic process – not a one-shot study – within an ongoing data management system. Program information is gathered and monitored during its operational life to determine if it is producing the desired results for disaggregated student populations and within various school contexts; or if it should be modified, revised, or discontinued. Staff training, support, additions, and other changes continue to occur during this evaluation process.

Once a summative evaluation of a program is completed, the district should have established a complete data management and program monitoring system. Such a system allows districts to acquire empirical estimates of the program's true costs and effects within actual school settings and to establish the relationship between time and use and use and effects. This system also allows the confirmative evaluations process to emerge and a program's enduring effects and benefits to be confirmed.

PART V: CONFIRMATIVE EVLAUATION

The rationale for golden rule ten is to answer the research question: How much is the program worth? The product of this research is a confirmative evaluation report with long-term program cost and benefit Information.

Golden Rule Number Ten

Produce a confirmative evaluation report after three years of program implementation.

The purpose of a confirmative evaluation is to verify the enduring effects and long-term benefits of a program. Unlike formative and summative evaluation, the program evaluation literature does not use the term 'confirmative' evaluation. Nor do schools and evaluators usually consider the issues to which it refers: identifying those changes that occur only after the passage of time and can be directly linked to participation in a program. While virtually every program, and indeed all educational efforts, is assumed to have some enduring impact, studies that actually identify and confirm these benefits are few.

Confirmative evaluation requires that several criteria be met. One is that the program has demonstrated its integrity in operational settings. This means that formative and summative evaluations have been carried out on the program and produced acceptable results. Second, there must be detailed assessment information available on the program participants; only then is there a basis for the evaluation of long-term effects. Finally, there must be instrumentation available or developed to assess the areas of possible effects, as well as other intervening factors that might have contributed to the long-term effects.

Confirmative evaluation also allows definitive information to be collected on many aspects of a program's actual costs. For example, schools at every level typically make frequent changes in their programs. Although most of these changes usually involve only using newer editions of the same program, they quite frequently involve changing entire programs. Since the vast majority of

school districts do not compile data on either their programs, the implementation of their programs, or the outcomes that their programs produce, even major program changes are made without any empirical rationale or careful calculations as to the actual costs involved in making such changes.

The data compiled on programs and students through the confirmative evaluation process provides the empirical answers to important cost information questions needed by district administrators and school boards, such as:

- Is the amount of time it typically takes teachers to learn how to use and implement the program still cost-effective?
- Is the amount of time most teachers need to prepare for each lesson/unit less than that with comparable programs?
- Is the amount of time needed for most students to learn the skills and concepts still adequate?
- Is the amount of supplementary/support materials needed still cost-beneficial?
- Is the program still continuing to produce the desired learner outcomes?
- Are the program effects worth its actual overall cost?

CONTRIBUTIONS TO THE FIELD OF EDUCATION

Use of this full R & D model has resulted in a number of significant contributions to the field of educational research and program evaluation. First, it created a new paradigm, not only for carrying out useful program evaluations, but also for analyzing and explaining both the failures of previous evaluation studies on schooling effects (e.g., the Coleman report and the Westinghouse Head Start studies) as well as the few notable successes (Novak and Musonda, 1991; Hanson and Farrell, 1995; Darmer, 1995).

Secondly, it shows the interrelationship between 1)-program evaluation and policy evaluation studies, and 2)-program evaluation and product development. To evaluate policy is to evaluate programs; to evaluate programs is to evaluate products. In simplest terms, it shows that the first step in any evaluation effort is to analyze these relationships to determine if there is indeed a product to evaluate and if further efforts are worth while.

Finally this R & D model provides a positive view of evaluation. If formative or evaluability analyses are not carried out, the results of further evaluations can be expected to be inconclusive in terms of both costs and benefits of the program or policy. Such traditional evaluations are not only wasteful, but also destructive. They add support to most practitioners' perceptions of program evaluation: that they are, at best, a major inconvenience and a waste of time and money that could be put to better use. At worst, they reinforce negative views of educational and evaluation reform efforts. This full R & D model allows evaluation to reclaim

it rightful role as a leader for promoting educational innovation and student leaning.

SUMMARY

Once a program has been through this R & D model of program evaluation, the relationships between time and program use and program use and effects can be firmly established, as well as it's validity in different settings and with different student populations. This process yields results that differ markedly from those typically found in program evaluation studies. In contrast to the prevalent simplistic notion that evaluations result in either retaining or rejecting a program, this approach provides insights into the costs and benefits, as well as where a program succeeds, where it fails, and WHY it succeeds or fails.

References

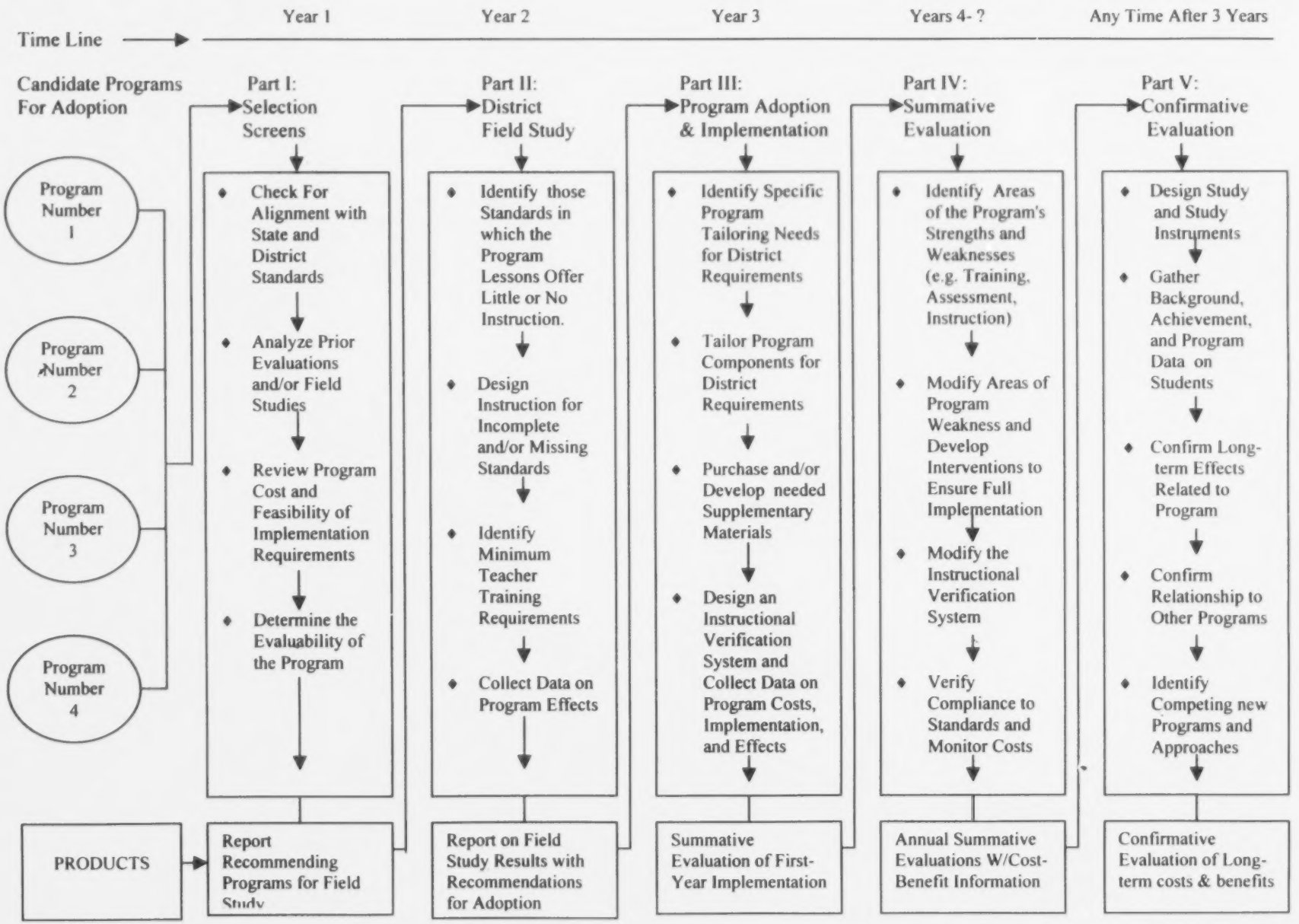
Darmer, M.A. (1995) Developing transfer and metacognition in educationally disadvantaged students: Effects of the Higher Order Thinking Skills (HOTS) Program. Unpublished dissertation, University of Arizona.

Hanson, R.A. & Farrell, D. (1995). The long-term effects of learning to read in kindergarten. Reading Research Quaterely, 30(4), 908-933.

Novak, J.D. & Musonda, D. (1991). A twelve-year longitudinal study of concept learning. American Educational Research Journal, 28(1), 117-153.

Wholey, J. S. (1976). Evaluation: Promise and Perfoemance. Washington, D.C.: Urban Institute.

Figure 1: TEMPLATE FOR ASSURING GAINS IN STUDENT LEARNING FROM SCHOOL PROGRAMS


Time Line
Year 1
Year 2
Year 3
Years 4- ?
Any Time After 3 Years
Candidate Programs For Adoption
Program Number 1
Program Number 2
Program Number 3
Program Number 4
Part I: Selection Screens
Check For Alignment with State and District Standards
Analyze Prior Evaluations and/or Field Studies
Review Program Cost and Feasibility of Implementation Requirements
Determine the Evaluability of the Program
Part II: District Field Study
Identify those Standards in which the Program Lessons Offer Little or No Instruction.
Design Instruction for Incomplete and/or Missing Standards
Identify Minimum Teacher Training Requirements
Collect Data on Program Effects
Part III: Program Adoption & Implementation
Identify Specific Program Tailoring Needs for District Requirements
Tailor Program Components for District Requirements
Purchase and/or Develop needed Supplementary Materials
Design an Instructional Verification System and Collect Data on Program Costs, Implementation, and Effects
Part IV: Summative Evaluation
Identify Areas of the Program's Strengths and Weaknesses (e.g. Training, Assessment, Instruction)
Modify Areas of Program Weakness and Develop Interventions to Ensure Full Implementation
Modify the Instructional Verification System
Verify Compliance to Standards and Monitor Costs
Part V: Confirmative Evaluation
Design Study and Study Instruments
Gather Background, Achievement, and Program Data on Students
Confirm Long-term Effects Related to Program
Confirm Relationship to Other Programs
Identify Competing new Programs and Approaches
PRODUCTS
Report Recommending Programs for Field Study
Report on Field Study Results with Recommendations for Adoption
Summative Evaluation of First-Year Implementation
Annual Summative Evaluations W/Cost-Benefit Information
Confirmative Evaluation of Long-term costs & benefits